

Durham Research Online

Deposited in DRO:

02 August 2013

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Deighton, J. and Tymms, P. and Vostanis, P. and Belsky, J. and Fonagy, P. and Brown, A. and Martin, A. and Patalay, P. and Wolpert, M. (2013) 'The development of a school-based measure of child mental health.', *Journal of psychoeducational assessment.*, 31 (3). pp. 247-257.

Further information on publisher's website:

<http://dx.doi.org/10.1177/0734282912465570>

Publisher's copyright statement:

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 License (<http://www.creativecommons.org/licenses/by-nc/3.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<http://www.uk.sagepub.com/aboutus/openaccess.htm>).

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

The Development of a School-Based Measure of Child Mental Health

Abstract

Early detection of child mental health problems in schools is critical for implementing strategies for prevention and intervention. The development of an effective measure of mental health and well-being for this context must be both empirically sound and practically feasible. This study reports the initial validation of a brief self-report measure for child mental health suitable for use with children as young as eight ("Me and My School" (M&MS)). After factor analysis, and studies of measurement invariance, two subscales emerged: emotional difficulties and behavioral difficulties. These two subscales were highly correlated with corresponding constructs of the Strengths and Difficulties Questionnaire (SDQ) and showed correlations with attainment, deprivation and educational needs similar to ones obtained between these demographic measures and the SDQ. Results suggest that this school-based self-report measure is psychometrically sound, and has the potential of contributing to school mental health surveys, evaluation of interventions, and recognition of mental health problems within schools.

Keywords: Mental health, self-report, child, school surveys, Me &My School

Early detection and assessment of child mental health problems is critical for implementing strategies for prevention and intervention (Weist, Rubin, Moore, Adelsheim, & Wrobel, 2007). Schools provide a key setting, common across children, for early recognition and intervention (Massey, Armstrong, Boroughs, Henson, & McCash, 2005; Roeser, Eccles, & Strobel, 1998). This focus on the school as a setting for intervention necessitates routine and widespread collection of information about child mental health and well-being (Levitt, Saka, Romanelli, & Hoagwood, 2007). The development of an effective measure of mental health and well-being for this context relies heavily on two key principles: it must be both empirically sound and practically feasible.

Two key elements contribute to the feasibility: who provides the information and compromises between breadth and brevity. In terms of who provides information, child self-report may be the most viable because there are practical constraints on the repeated use of teachers' time and the ability to gain sufficient response rates from parents. Furthermore, both parents and teachers tend to be less accurate in their assessment of emotional than behavioral difficulties (e.g., Tremblay, Vitaro, Gagnon, Piche, & Royer, 1992).

A systematic review identified 113 child mental health assessment measures (Wolpert et al., 2008), but no self-report assessments of general mental health (emotional and behavioral) were identified for children below the age of 11 that could be used as a brief community-based screening measure.

There are mixed views in relation to children's self-ratings of well-being, with particular concern in the literature that younger children may not be reliable informants about their own mental health (Roy, Veenstra, & Clench-Aas, 2008). However, analysis of extant measures (Wolpert et al., 2008) indicated that a likely reason for the limitations of self-report measures for this age group may be that measures developed for older children use inappropriate language for younger children or those with low levels of reading and/or language skills. Evidence suggests that children as young as six can reliably self-report if an age appropriate measure is used (Riley, 2004), especially where measures are developed specifically for this age group and used in community settings (Muris, Meesters, Eijkelenboom, & Vincken, 2004). An under-utilized area in the administration of these types of measures to children, which may facilitate use with younger children, is the use of computers with audio-feeds to reach those whose reading skills are poor.

In order to capture community-wide prevalence and trends in common mental health and well-being difficulties, measures need to be sufficiently brief to allow for routine use, yet be sufficiently wide-ranging to cover broad categories of the most common psychological difficulties. The most commonly reported mental health difficulties are either behavioral or emotional (Green, McGinnity, Meltzer, Ford, & Goodman, 2005; Levitt et al., 2007). These are also the domains most commonly covered by existing measures of mental health and well-being such as the Achenbach System of Behaviour Assessment (ASEBA; Achenbach & Rescorla, 2001) and the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997). In terms of empirical soundness, the scores derived from a range of measures of child mental health and well-

being have shown good reliability in both clinical and community settings and there is evidence for the possibility of valid interpretation (e.g., SDQ , Goodman, 1997; ASEBA, Achenbach & Rescorla, 2001). Some of these have even adopted approaches to validation using new, more advanced psychometric techniques, such as Item Response Theory (IRT; Chorpita et al., 2010). However, few of these measures have been developed specifically for the school setting and none of these existing instruments, to the authors' knowledge, offer a self-report measure for children under the age of 11 that is both brief and free.

Aims

The aim of this paper is to present the development and initial validation of a brief self-report measure for child mental health and well-being, as indexed by emotional and behavioral difficulties, suitable for use with children as young as eight years old. Specifically it describes the constructs underlying responses to the measure, score reliability, studies of measurement invariance in regard to several subgroups, construct validity, as well as the development of clinical cut-off scores based on an established mental health measure.

Materials and Methods

Sample

Data were collected in 2008 from pupils who attended state schools in 25 local areas across England. The analyses reported are based on surveys completed by 9,814 pupils aged 8-9 years (school year 4; 51.4% male) from 311 primary schools and 9,881 pupils aged 11-12 years (school year 7; 49.8% male) from 82 secondary schools. The initial sample was not drawn to be representative of all school children in England; it was

based on each local area's own selection of schools to be involved in a wider national program of child mental health provision in schools (DCSF, 2008).

The average academic attainment for children in the sample was slightly lower than the national average (primary schools: national average = 15.30, sample average = 14.84, SD= 3.63; secondary schools: national average = 27.70, sample average = 27.24, SD=4.52). They also had a slightly elevated level of deprivation, as measured by the Income Deprivation Affecting Children Index (IDACI) scores, compared to the national figures (primary schools: national average = 0.24, sample average = 0.291, SD=0.19 ; secondary schools: national average = 0.22, sample average = 0.28, SD= 0.19). Finally, there were similar proportions of children belonging to the 'White British' ethnic category compared to national figures (primary schools: national percentage = 73.8%, sample percentage = 74.1%; secondary schools: national percentage = 77.3%, sample percentage = 76.8%; Department for Education, 2010).

Procedure

Children completed questionnaires using a secure online system during their usual school day with parent consent. Teachers explained to participating children what the questionnaire was about, the confidentiality of their answers and their right to decline participation. The online system was designed to be easy to read and child-friendly; the font size was large and the instructions and individual questions were presented slowly to allow for less accomplished readers. For younger children, recorded spoken accompaniment for all instructions, questionnaire items and response options was also provided. Parents were also invited to complete a questionnaire about their child.

Measures

Me & My School (M&MS). The M&MS measure was developed to cover two broad domains: emotional difficulties and behavioral difficulties. Based on a review of outcome measures (Wolpert et al, 2008) and an analysis of key concepts covered by the emotional and behavioural scales of other measures a large pool of items was generated keeping in mind lower reading age and usability by younger age groups. This larger pool of items was piloted in focus groups with children to establish which terms and concepts younger children used and understood. This process resulted in an initial pool of 24 items included in the online questionnaire (see Table 1). Items consisted of short behavioral statements to which children responded using the response options “never”, “sometimes” or “always”. The items were converted to an online survey, designed to be visually clear and appealing to children.

The Strengths and Difficulties Questionnaire (SDQ). The SDQ (Goodman, 1997) is a widely used and well validated mental health measure with subscales including behavioral and emotional problems and validated cut-offs indicating thresholds distinguishing children who are likely to require clinical intervention and those who are not (Goodman, Meltzer & Bailey, 1998). As this measure is for children aged 11 and over, it was only completed by children who were 11-12 years in the current sample.

Additional variables. Additional variables were drawn from the English National Pupil Dataset, a government dataset holding records on all pupils in England.

Academic attainment was recorded based on children’s performance in nationally mandated assessments and the scores achieved for the most recent assessment were used

for each age group. Scores were based on each child's attainment in three core subject areas (English, mathematics and science), high scores representing high attainment.

Deprivation scores were defined using the deprivation score associated with the child's home postcode, known as the Income Deprivation Affecting Children Index (IDACI, Mc Lennan, Barnes, Noble, Davies, Garratt & Dibben, 2011). IDACI scores are on a metric between 0 and 1 with higher scores reflecting higher levels of deprivation.

Special Educational Needs (SEN) were based on the school's assignment of a child to a level of special educational needs. These were coded on a four-point scale as follows: 0 = No statement of SEN, 1 = School Action, 2 = School Action Plus, and 3 = Statement of SEN. Therefore, the higher the score, the greater the educational needs.

Analyses

Analyses were carried out in several stages. First, Exploratory Factor Analysis (EFA) was performed on a randomly selected half of the total sample, in order to determine which items constituted coherent subscales of emotional and behavioral difficulties. The items with most salient factor loadings and least cross-loadings were selected and Confirmatory Factor Analysis (CFA) confirming this item assignment was performed on the other half of the sample. Second, Differential Item Functioning (DIF) analyses were used to confirm whether the items were suitable for use across a range of demographic groupings. Third, Cronbach's alpha (α) was used to assess internal consistency of the derived scores based on the retained items. Fourth, validity of the scores was investigated by correlating the emotional and behavioral subscales with pre-existing mental health scales and other theoretically related constructs, such as academic attainment and deprivation. Finally, cut-offs indicating those at high risk of mental

health problems were established for the new measure by equipercentile equating to cut offs on a pre-existing well validated instrument (SDQ).

Results

Factor Structure

EFA and Parallel analysis (Hoyle & Duvall, 2004) carried out on the polychoric correlations of the 3-category items (Muthén & Kaplan, 1985) on half of the sample (N=9,837) suggested presence of three factors (first four eigenvalues were 6.87, 2.86, 1.81 and 1.02). Goodness of fit of the three-factor solution was good (CFI = 0.955; TLI = 0.940; RMSEA = 0.044; SRMR = 0.035), and substantially better than a two-factor solution (CFI = 0.898; TLI = 0.877; RMSEA = 0.064; SRMR = 0.058). An oblique rotation of the factors yielded a solution presented in Table 1. The items designed to measure behavioral difficulties clustered together yielding high loadings (Factor III). The emotional difficulties items also largely loaded on a single factor (Factor I), although some were more clearly linked to the second factor. The second factor (Factor II) was made up of a small set of items related to social aspects of child's life such as friendships, which were initially developed to form part of the emotional difficulties subscale. The correlations between factors were positive but low, confirming that they represented related but conceptually distinct constructs.

[Insert Table 1 about here]

Confirmatory Factor Analysis was carried out on the second half of the sample (N = 9,858). We tested a model with two correlated factors indicated by the items identified in EFA as belonging to the emotional and behavioral difficulties (bold items under Factors I and III in Table 1). The items were hypothesized to indicate only one factor

(independent clusters structure; McDonald, 1999). The model's goodness of fit (CFI = 0.934; TLI = 0.924; RMSEA = 0.060) was acceptable (Hu & Bentler, 1999). The correlation between the two latent factors was 0.42 ($p < .001$). The standardized factor loadings from this model are presented in Table 1.

Differential Item Functioning (DIF)

DIF analysis has been used to inform the selection of suitable items which operate equivalently across a range of different subgroups of school children. Five grouping criteria were examined: gender, special education needs (SEN), whether English was the child's second language (EAL), whether the child received free school meals (FSM), and whether the child was in care. Girls (49.5%), children with SEN (1.9% statemented), non-native English speakers (17.5%), children receiving FSM (21%), and children in care (0.5%) were the focus of these investigations (formed the *focal* groups in the DIF analyses). DIF analyses compare the item endorsement rates in the focal group compared to the *reference* group (for example, children in care versus all other children), conditioning on the test score. An item is said to display DIF if children with the same test score but belonging to different groups have different probabilities of endorsing the item.

The statistical approach taken was the Liu-Agresti common log odds ratio (L-A LOR; Liu & Agresti, 1996), a non-parametric Mantel and Haenszel-type estimator. The L-A LOR relies on the log odds ratio of one group selecting a particular response option relative to the other group, stratified by overall level of the measured construct. Both constructs (emotional and behavioral difficulties) were examined using this method. Table 2 shows results of DIF analyses performed with DIFAS 5.0 (Penfield, 2005).

Positive L-A LOR values indicate the item is more difficult to endorse for the focal group; negative values indicate that the item is easier to endorse, given the same level of the construct. L-A LOR values are printed in the table only if: 1) they are statistically significant at the 0.05 level; and 2) LOR is at least moderate in size ($|L-A LOR| > 0.43$; Penfield, 2007).

[Insert Table 2 about here]

Analysis indicated that “Other children tease me” and “I bully others” were the most problematic items, both in terms of the magnitude and the number of groupings showing DIF. Gender differences were identified for some emotional items but were not considered sufficient alone to warrant removal of items since gender differences in presentation of emotional distress are well known (Rutter, Caspi, & Moffitt, 2003). Boys, those with SEN and those in care were more likely to agree that other children teased them than were others with similar levels of emotional difficulties. Also, those with SEN, EAL and in care were more likely to agree that they bullied others than were others with similar levels of behavioral difficulties. Based on this, both of these items were removed from the final scale, resulting in a ten-item scale of emotional difficulties and a six-item scale of behavioral difficulties (Table 2).

Internal Consistency

Cronbach’s alpha for the two resulting scales was good (behavioral difficulties: $\alpha = 0.78$ and 0.80 for Years 4 and 7 respectively; emotional difficulties: $\alpha = 0.72$ and 0.77). In comparison, the alpha for the corresponding SDQ subscales in the Year 7 sample was slightly lower (emotional symptoms: $\alpha = 0.72$; conduct problems: $\alpha = 0.68$).

Validity

The construct validity of the scores derived from the M&MS measure was assessed by considering the convergent and divergent validity of each subscale against the relevant subscales of the SDQ. Additional validity evidence was gained by considering behavioral and emotional difficulties with theoretically related constructs of academic attainment, deprivation and SEN.

Construct validity. As the self-report SDQ can only be used for children aged 11 years or over, correlations between the M&MS scales and the self-report SDQ subscales could only be computed for the older age group (Table 3). In this group, correlations between the emotional and behavioral M&MS scales, and the corresponding self-report SDQ subscales were high ($r = .67, p < .001$; $r = .70, p < .001$). It is also notable that the correlations between the M&MS subscales and the non-corresponding self-report SDQ subscale were much lower ($r = .22, p < .001$; $r = .24, p < .001$), suggesting good discriminant validity.

[Insert Table 3 about here]

Correlations between M&MS scales and the corresponding parent report SDQ subscales (Tables 3 and 4) were lower but remained statistically significant and were comparable to correlations observed between the SDQ self-report and parent-report, where these data were available (i.e., for the older group).

[Insert Table 4 about here]

Additional validity evidence. Three variables documented to have a moderately strong relationship with emotional and behavioral difficulties were considered: academic attainment, extent of SEN, and deprivation (Masten et al., 2005; Reijneveld et al., 2010;

Meltzer, Gatward, Goodman, & Ford, 2003). Consistent with the literature, each of these variables correlated more highly with the two behavioral scales than the emotional scales (see Table 5). In each instance higher levels of emotional or behavioral difficulties were associated with lower academic attainment, greater extent of SEN and higher deprivation. Correlations between these variables and the M&MS subscales were statistically significant due to large samples used, but small in magnitude. These relationships are, however, consistent with correlations observed for the SDQ subscales (see Table 5).

[Insert Table 5 about here]

Establishing Clinical Cut-offs

Clinical cut-offs were established for the M&MS measure against the already established cut-offs for the SDQ using equipercentile equating (Kolen & Brennan, 2004). For the SDQ emotional symptoms subscale, a score of six is borderline and scores seven and above are clinically significant (Goodman, Meltzer & Bailey, 1998). In our SDQ sample of Year 7 pupils, these scores corresponded to percentile ranks¹ 91.9 and 95.7 respectively. These percentile ranks translate into cut-off scores of 10-11 for borderline and 12 and above for clinically significant for the M&MS emotional difficulties subscale. For the SDQ conduct problems subscale the recommended cut-offs of four (borderline) and five (clinically significant) corresponded to percentile ranks 85.5 and 91.9 in our SDQ sample of Year 7 pupils, translating into cut-off scores of six for borderline and seven and above for high risk for the M&MS behavioral difficulties subscale.

The M&MS high-risk cut-off identified 12 per cent of younger children (Year 4) and 12.4 per cent of older children (Year 7) in the sample as having behavioral

¹ Percentiles indicate the percentage of children obtaining this score or lower on the scale.

difficulties (i.e., obtained a score of seven or above), and 10.2 per cent of younger children and 5.8 per cent of older children in the sample as having emotional difficulties (i.e., obtained a score of 12 or above).

Discussion

Initial validation suggests this is a psychometrically sound brief self-report measure of mental health and well-being for young children, functioning equivalently across described groups.

The final M&MS subscales showed strong relationships with the self-report SDQ, suggesting that they measure similar underlying constructs. The discrepancy between the high magnitude of the conceptually similar subscales and the relatively low magnitude of the conceptually distinct subscales provides support for the construct validity of the scores (John & Benet-Martinez, 2000). Correlations between M&MS subscales and measures of academic attainment, SEN and deprivation were low but statistically significant and consistent with correlations observed between these measures and the SDQ subscales.

In terms of limitations, it should be noted that the current analyses were based on a web-based delivery of measures, responses on paper versions are still to be tested. A limitation with regards to establishing clinical cut-offs is that the derived scores have not yet been validated with clinical populations. Stability of scores (test-retest reliability) will also need to be established. All of these areas will be explored in future research.

Whilst further work is needed, the M&MS appears to fill a number of gaps. It was designed specifically for use in schools and allows even young children to report their

own experience of mental health and well-being. It has potential for use as a broad screening tool to aid detection of child mental health problems and evaluation of school-based child mental health interventions.

Acknowledgements

We would like to thank other members of the research group: Professor Neil Humphrey, Professor Norah Frederickson, Pam Meadows, Professor Anthony Fielding, Dr Rob Coe, Mike Cuthbertson, Neville Hallam, Dr John Little, Dr Andrew Lyth, Professor Sir Michael Rutter, Professor Bette Chambers and Professor Alastair Leyland. We would like to thank the English Department for Children Schools and Families (now the Department for Education) for funding the research.

References

- Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profile: an integrated system of multi-informant assessment*. Burlington, Vt.: ASEBA
- Chorpita, B. F., Reise, S., Weisz, J. R., Grubbs, K., Becker, K. D., & Krull, J. L. (2010). Evaluation of the Brief Problem Checklist: child and caregiver interviews to measure clinical progress. *Journal of Consulting and Clinical Psychology*, 78(4), 526-536. doi:10.1037/a0019602
- McLennan, D., Barnes, H., Noble, M., Davies, J., Garatt, E., & Dibben, C., (2011). The English Indices of Deprivation 2010: Technical Report. Department for Communities and Local Government. Retrieved from <http://www.communities.gov.uk/communities/research/indicesdeprivation/deprivation10/>
- DCSF. (2008). *Targeted mental health in schools project : using the evidence to inform your approach: a practical guide for head teachers and commissioners*. Nottingham: Dept. for Children, Schools and Families
- Department for Education. (2010). *Schools, Pupils and their Characteristics: January 2010*. . Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/s000925/sfr09-2010.pdf>.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry*, 38(5), 581-586

- Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. *European Child Adolescent Psychiatry*, 7(3), 125-130
- Green, H. (2005). *Mental health of children and young people in Great Britain, 2004*. Basingstoke: Palgrave Macmillan
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. doi: 10.1080/10705519909540118
- John, O. P., & Benet-Martínez, V. (2000). Measurement, scale construction, and reliability. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 339-369). New York: Cambridge University Press.
- Kolen, M. J., Brennan, R. L., & Kolen, M. J. T. e. (2004). *Test equating, scaling, and linking: methods and practices* (2nd ed. ed.). New York: Springer
- Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology*, 45(2), 163-191. doi:dx.doi.org/10.1016/j.jsp.2006.11.005
- Liu, I. M., & Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratios with a stratified ordinal response. *Biometrics*, 52(4), 1223-1234
- Hoyle, R. H. and Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In D. Kaplan (Ed.): *The Sage handbook of quantitative methodology for the social sciences*. Thousand Oaks, CA: Sage.

- Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi: 10.1080/10705519909540118
- Massey, O. T., Armstrong, K., Boroughs, M., Henson, K., & McCash, L. (2005). Mental health services in schools: A qualitative analysis of challenges to implementation, operation, and sustainability. *Psychology in the Schools*, 42(4), 361-372. doi:10.1002/pits.20063
- Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradovic, J., Riley, J. R., et al. (2005). Developmental cascades: linking academic achievement and externalizing and internalizing symptoms over 20 years. *Dev Psychol*, 41(5), 733-746. doi:10.1037/0012-1649.41.5.733
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.J.: L. Erlbaum Associates.
- Meltzer, H., Gatward, R., Goodman, R., & Ford, T. (2003). Mental health of children and adolescents in Great Britain. *Int Rev Psychiatry*, 15(1-2), 185-187. doi:10.1080/0954026021000046155
- Muris, P., Meesters, C., Eijkelenboom, A., & Vincken, M. (2004). The self-report version of the Strengths and Difficulties Questionnaire: its psychometric properties in 8- to 13-year-old non-clinical children. *British Journal of Clinical Psychology*, 43(Pt 4), 437-448. doi:10.1348/0144665042388982

- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189. doi:10.1111/j.2044-8317.1985.tb00832.x
- Penfield, R. (2005). DIFAS: Differential Item Functioning Analysis System. *Applied Psychological Measurement*, 29(2), 150-151. doi:10.1177/0146621603260686
- Penfield, R. D. (2007). An Approach for Categorizing DIF in Polytomous Items. *Applied Measurement in Education*, 20(3), 335-355. doi:10.1080/08957340701431435
- Reijneveld, S. A., Veenstra, R., de Winter, A. F., Verhulst, F. C., Ormel, J., & de Meer, G. (2010). Area deprivation affects behavioral problems of young adolescents in mixed urban and rural areas: the TRAILS study. *Journal of Adolescent Health*, 46(2), 189-196. doi:10.1016/j.jadohealth.2009.06.004
- Riley, A. W. (2004). Evidence that school-age children can self-report on their health. *Ambul Pediatr*, 4(4 Suppl), 371-376. doi:10.1367/A03-178R.1
- Roeser, R. W., Eccles, J. S., & Strobel, K. R. (1998). Linking the study of schooling and mental health: Selected issues and empirical illustrations at. *Educational Psychologist*, 33(4), 153-176. doi:10.1207/s15326985ep3304_2
- Rutter, M., Caspi, A., & Moffitt, T. E. (2003). Using sex differences in psychopathology to study causal mechanisms: unifying issues and research strategies. *Journal of Child Psychology and Psychiatry*, 44(8), 1092-1115. doi: 10.1111/1469-7610.00194
- Tremblay, R. E., Vitaro, F., Gagnon, C., Piché, C., & Royer, N. (1992). A prosocial scale for the preschool behaviour questionnaire: concurrent and predictive correlates.

International Journal of Behavioral Development, 15(2), 227-245.

doi:10.1177/016502549201500204

Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry*, 49(12), 1304-1312. doi:10.1111/j.1469-7610.2008.01942.x

Weist, M. D., Rubin, M., Moore, E., Adelsheim, S., & Wrobel, G. (2007). Mental Health Screening in Schools. *Journal of School Health*, 77(2), 53-58. doi:10.1111/j.1746-1561.2007.00167.x

Wolpert, M., Aitken, J., Syrad, H., Munroe, M., Saddington, C., Trustam, E., et al. (2008). *Review and recommendations for national policy for England for the use of mental health outcome measures with children and young people*. London: Department for Children, Schools and Families. Retrieved from: <http://www.ucl.ac.uk/clinical-psychology/EBPU/publications/reports.php>

Table 1

EFA Rotated Loadings, CFA Standardized Loadings, and Correlations Between Factors

Original assignment	Item	EFA Factors			CFA Factors	
		I	II	III	I	II
Emotional difficulties	I feel happy	.15	.43	.18		
	I feel lonely	.48	.29	-.05	.62	
	I am unhappy	.41	.32	.10	.59	
	I like the way I look	.13	.37	-.02		
	Nobody likes me	.38	.32	.03	.54	
	I enjoy break times	.12	.43	.01		
	I enjoy playing with friends	.05	.65	.11		
	I cry a lot	.61	.02	-.02	.55	
	Other children tease me	.51	.20	.11	.65	
	I worry when I am at school	.65	.13	.01	.68	
	I worry a lot	.63	.06	-.07	.58	
	I have problems sleeping	.58	-.11	.14	.56	
	I have lots of friends	.30	.61	-.05	.56	
	I wake up in the night	.50	-.16	.19	.51	
	I am shy	.39	.02	-.13	.30	
	I feel scared	.70	-.01	-.07	.63	
	I enjoy being with other children	-.04	.60	.18		
Behavioral difficulties	I get very angry	.22	-.01	.73		.82
	I lose my temper	.17	-.01	.79		.85
	I bully others	-.02	.16	.67		.69
	I do things to hurt people	.00	.14	.75		.75
	I am calm	-.04	.31	-.49		-.80
	I hit out when I am angry	.05	.05	.78		.62
	I break things on purpose	.01	.14	.59		.82
Factor correlations	Factor 1	1			1	
	Factor 2	.30	1		.42	1
	Factor 3	.24	.13	1		

Note. Both EFA and CFA were performed on polychoric correlations using diagonally weighted estimator. Rotation method in EFA: Geomin. All loadings in CFA and factor correlations in both EFA and CFA are significant at $p < 0.000$.

Table 2

Differential Item Functioning for the M&MS Scales

Item	Gender (focal=girls)	SEN (focal=yes)	EAL (focal=yes)	FSM (focal=yes)	In Care (focal=yes)	Item to be used in final scales?
I feel lonely						Yes
I am unhappy						Yes
Nobody likes me						Yes
I cry a lot	-.47					Yes
Other children tease me	.62	-.48			-.43	No
I worry when I am at school						Yes
I worry a lot						Yes
I have problems sleeping	.52					Yes
I wake up in the night						Yes
I am shy	-0.52					Yes
I feel scared	-0.47					Yes
I get very angry						Yes
I lose my temper						Yes
I bully others		-0.69	-0.60		-0.63	No
I do things to hurt people						Yes
I am calm		-0.47				Yes
I hit out when I am angry	0.49					Yes
I break things on purpose						Yes

Note. EAL = English as an additional Language; FSM = Free School Meals; SEN = Special Educational Needs.

* Very few cases are identified as 'In Care'; results presented are significant but might be unstable.

Table 3

Correlations Between M&MS and SDQ Subscales for Children Aged 11-12 (Year 7)

Variable	1.	2.	3.	4.	5.
1. Emotional difficulties M&MS	-				
2. Behavioral difficulties M&MS	.29**	-			
3. SDQ self-report emotional symptoms	.67**	.24**	-		
4. SDQ self-report conduct problems	.22**	.70**	.26**	-	
5. SDQ parent-report emotional symptoms	.33**	.09*	.37**	.09*	-
6. SDQ parent-report conduct problems	.15**	.29**	.15**	.36**	.36**

Note. * $p < .05$; ** $p < .01$; sample size for child-parent/parent-parent assessments varied from 579 to 655, sample size for child-child assessments varied from 8,138 to 9,324

Table 4

Correlations between M&MS and SDQ Subscales for children aged 8-9 (Year 4)

Variable	1.	2.	3.	4.
1. Emotional difficulties M&MS	-			
2. Behavioral difficulties M&MS	.35**	-		
3. SDQ parent report emotional symptoms	.17**	.10*	-	
4. SDQ parent report conduct problems	.06	.31**	.38**	-

Note. * $p < .05$; ** $p < .01$; sample size for child-parents/parent-parent assessments varied from 609 to 675, sample size for child-child assessments was 8,300

Table 5

Correlations between M&MS and Conceptually Related Variables

Year	Year 4			Year 7		
Variable	Academic attainment	SEN	Deprivation	Academic attainment	SEN	Deprivation
M&MS emotional difficulties	-.13**	.09**	.05**	-.12**	.10**	.03**
SDQ emotional symptoms	-	-	-	-.14**	.10**	.05**
M&MS behavioral difficulties	-.19**	.17**	.11**	-.19**	.16**	.13**
SDQ conduct problems	-	-	-	-.25**	.18**	.13**

Note. *p<.05; **p<.01; sample size varied from 8,476 to 9,246